PCAS 17 (2014/2015)

Supervised Project Report
(ANTA604)

DERIVING ANTARCTIC SEA ICE THICKNESS FROM REMOTE SENSING USING ARTIFICAL NEURAL NETWORKS AND GENETIC ALGORITHM

Benjamin J. E. Schroeter

Student ID: 53667376 (bsc41)

Word count: 5876

Abstract/executive summary (ca. 200 words):

This paper investigates the applicability of the Artificial Neural Network (ANN) learning paradigm in deriving sea ice thickness in the Antarctic using a combination of physical measurements, Landsat-8 satellite imagery, and Synthetic Aperture Radar (SAR) imagery as training input, validated against physical thickness measurements. The paper discusses the complexities of sea ice and the difficulties faced in calculating thickness before detailing the current literature applying ANNs to predicting sea ice characteristics. Methodology is described in terms of an overview of ANNs and the optimisations applied in this study, as well as the geospatial processing and data manipulation undertaken to produce the training and test data sets from sparse point measurements.

Results indicate that physical field measurements are the principle contributor to model performance, outperforming the Landsat-8 model and increasing performance in a combined approach. Error for thickness prediction varied between models: ±14cm (physical measurements), ±23cm (Landsat-8), and ±21cm (combined Landsat-8 and physical measurements). The paper concludes with a number of suggested improvements to the model that future studies should consider as well as calling for further validation of the model through additional, actual field measurements over manufactured data points produced through interpolation.

DERIVING ANTARCTIC SEA ICE THICKNESS FROM REMOTE SENSING USING ARTIFICAL NEURAL NETWORKS AND GENETIC ALGORITHM

BENJAMIN J. E. SCHROETER

SUPERVISOR: DR WOLFGANG RACK

POSTGRADUATE CERTIFICATE IN ANTARCTIC STUDIES
CANTERBURY UNIVERSITY, CHRISTCHURCH

ABSTRACT

This paper investigates the applicability of the Artificial Neural Network (ANN) learning paradigm in deriving sea ice thickness in the Antarctic using a combination of physical measurements, Landsat-8 satellite imagery, and Synthetic Aperture Radar (SAR) imagery as training input, validated against physical thickness measurements. The paper discusses the complexities of sea ice and the difficulties faced in calculating thickness, before detailing the current literature applying ANNs to predicting sea ice characteristics. Methodology is described in terms of an overview of ANNs and the optimisations applied in this study, as well as the geospatial processing and data manipulation undertaken to produce the training and test data sets from sparse point measurements.

Results indicate that physical field measurements are the principle contributor to model performance, outperforming the Landsat-8 model and increasing performance in a combined approach. Error for thickness prediction varied between models: ±14cm (physical measurements), ±23cm (Landsat-8), and ±21cm (combined Landsat-8 and physical measurements). The paper concludes with a number of suggested improvements to the model that future studies should consider as well as calling for further validation of the model through additional, actual field measurements over manufactured data points produced through interpolation.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

INTRODUCTION

Measuring the thickness of Antarctic sea ice has historically been one of the more difficult sea ice characteristics to map due to its complex dynamics and characteristics (Lemke et al., 2007). A large proportion of sea ice in the Southern Ocean is first-year ice of approximately 1-2 metres thick (Wadhams, 2000). However, thickness can be significantly reduced by divergence, or increased through convergence (Lemke et al., 2007). For example, in low convergence scenarios where ice floes in the ocean collide, they can "raft" and slide beneath each other, doubling their thickness. Closer to the coast, the pressure of wind driving sea ice against land mass can cause strong convergence, leading to the formation of ridges. Basal freezing also thickens sea ice during the winter, acting to insulate the ocean and prevent heat loss into the atmosphere (Lemke et al., 2007). However, it is not only the thickness of the ice itself that is of interest; it is also the thickness of the overlying snow, which significantly alters the geometry and insulating dynamics of the sea ice (Lemke et al., 2007; Xie et al., 2011).

Historically, measurements were conducted by moored sonar instruments detecting *draft,* the submerged portion of the ice floe, which were limited to short-timescales (Lemke et al., 2007). Even where longer-term datasets from submarine or ship-based observations exist for the Arctic, the general spatial and temporal paucity of sea ice thickness data limits observations of long-term trends (Solomon et al., 2007). During the last decade, significant progress has been made in the development of ship-based observations and the use of satellite instrumentation to retrieve data of sea ice characteristics (Solomon et al., 2007).

Different sea ice characteristics are retrieved by different remote sensing instruments; in the case of sea ice thickness, recent years have seen the use of radar and laser altimeters to conduct large-scale annual surveys in both the Arctic and Antarctic (Kurtz et al., 2013). However, currently, only signals emitted from the top few tens of centimetres can be detected and retrieved by satellite sensors, leaving the lower surface unmeasured (Kwok, 2011). Freeboard measurements, the amount of sea ice and snow above the sea surface, are derived from altimetry measurements of reflectivity, waveform, and elevation parameters, and calculated using the assumption of hydrostatic equilibrium to give estimates of sea ice thickness (Kwok, 2011). While these large-scale surveys of satellite altimetry have yielded reasonable estimates of sea ice thickness in the Arctic, it is considerably more difficult in the Antarctic due to the difference in characteristics between Antarctic and Arctic sea ice (Xie et al., 2011). For example, due to lower levels of freshwater input in the Southern Ocean, and the lower levels of brine rejection in first year ice, salinity in the surface ocean waters and sea ice is higher in Antarctica than in the Arctic (Wadhams, 2000). Antarctic sea ice is also generally quite thin and warm compared to Arctic sea ice, and more mobile (Solomon et al., 2007). Lastly, measuring freeboard in Antarctica is made difficult due to snow coverage and metamorphism (Kwok, 2011). High precipitation levels over Antarctic sea ice lead to thicker snow cover than in the Arctic, pressing down on the sea ice and

driving floes beneath sea level where the snow becomes saturated with seawater (Solomon et al., 2007). It then quickly refreezes, forming a new ice type called 'snow-ice' (Wadhams, 2000). This results in a significantly more complex geometry than Arctic sea ice, with multi-layer geometric systems across individual ice floes problematic for algorithms accounting for buoyancy theory (Xie et al., 2011). The rapid metamorphism of snow-ice production has led to calls for sea ice models to incorporate not only regional and seasonal differences in snow and ice thickness between the Arctic and Antarctic, but also time-dependent changes in snow parameters, as the rate of change has significant impacts on the thermal conductivity of the sea ice floe (Massom et al., 2001).

The complexities of sea ice formation have encouraged researchers to seek new ways to understand and model the non-linearity of physical and optical characteristics of the phenomenon. This has lead to an interest in the applicability of Machine Learning techniques, which have no physical basis, to attempt to expose the implicit structure or relationship between sea ice characteristics that are not yet fully understood in the physical sciences. One such technique that has garnered interest in the polar community is the Artificial Neural Network (ANN), which has application in a number of key problem areas: classification, regression, and function approximation (Engelbrecht, 2007), particularly functions that are highly non-linear (El-Rabbany et al., 2002).

ANNs have been employed for the purposes of deriving a number of sea ice characteristics such as sea ice extent edge (Alhumaidi et al., 1997), concentration (Belchansky et al., 2003; El-Diasty, 2003), sea ice type (such as first-year or multi-year) (El-Rabbany et al., 2002), and sea ice thickness (Lin & Yang, 2012; Yap et al., 2009). These studies have made use of different sources of physical and remotely-sensed observations such as visual classifications (El-Rabbany et al., 2002), Radar/SAR (Alhumaidi et al., 1997), Passive Microwave (Belchansky et al., 2003) and optical satellite imagery as training data for the model, using either archival or physical measurements as "ground-truth" model validation.

These studies have experienced varying levels of success, from model underperformance due to domain size (El-Rabbany et al., 2002), to classification accuracy of up to 98% (Alhumaidi et al., 1997) in problem domains across the globe. However, there are few literature examples describing the application of the ANN to Antarctic sea ice thickness derivation.

This study will investigate the applicability of Artificial Neural Networks in deriving Antarctic sea ice thickness from a combination of physical measurements made by Rack (2013/14) and remotely sensed observations from satellite using Landsat-8 and Synthetic Aperture Radar (SAR) imagery. Using a combination of geospatial and data interpolation techniques, a training data set is generated and used to train the model.

METHOD

The methodology employed for this study consists of two principle components: the development and optimisation of the Artificial Neural Network algorithm (written in Python), and the generation, selection and cleansing of data for training and validation using the Quantum GIS (QGIS) software package and custom Python code.

ARTIFICIAL NEURAL NETWORKS

The origin of the Artificial Neural Network (ANN) can be traced back to Rosenblatt's "Perceptron" (Rosenblatt, 1958), whereby the neurone of the brain were approximated via an algorithm that mimicked the activation of neural synapses, building on the threshold logic research of McCulloch and Pitts (1943). The perceptron operates on the premise of feeding a series of input signals (values) x_i into the model. These signals are then multiplied by weights w_i and passed through a transfer function (typically standard summation), the output of which would then in turn be fed into an activation function producing an output signal for the perceptron (Figure 1).

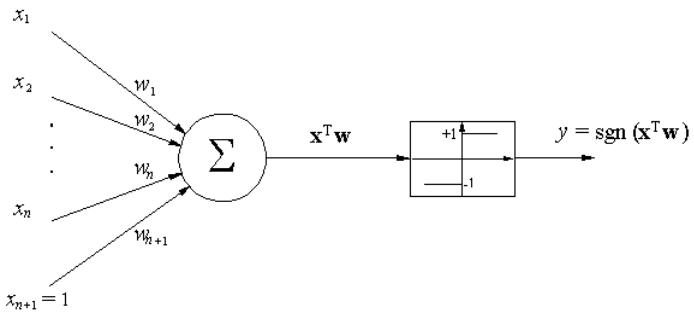

Figure 1 - A schematic of Rosenblatt's perceptron (Hassoun, 1995)

Activation functions come in many forms, the simplest of which is the step function. This function simply takes the output from the transfer function and produces a signal of 0 or 1 depending on the position of the step (the bias). Other forms of activation functions include the Sign and Sigmoid (Figure 2), as well as linear activation or mathematical variations of the Sigmoid, which effectively change the slope of the activation (Bedo, 2014a).

For this study, we used the sigmoid activation function as modelled by the hyperbolic tangent, $\tanh(x)$, as it effectively models continuous real values (such as is required by ice thickness prediction).

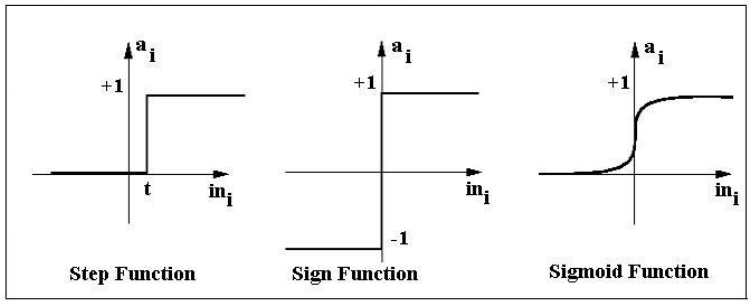

Figure 2 - Examples of common activation functions for ANNs (Kendall, 2001)

Research into ANNs stalled for a number of years when Minsky and Papert (1969) released a paper that demonstrated that Rosenblatt's Perceptron could not learn even the most basic of computational reasoning routines, the Exclusive Or (XOR) function. It wasn't until Rumelhart et al. (1988) discovered the back-propagation algorithm that research began anew, and the ANN was positioned as one of the most advanced non-linear predictive models of the modern day, commanding significant interest from computer science researchers (Engelbrecht, 2007).

Contemporary ANNs function in two distinct stages: Forward-Feeding, and Back-Propagating.

FEED-FORWARD

The forward-fed ANN, or multi-layered perceptron (MLP) is described in terms of neurone and hidden layers. With the exception of the input layer (the instance), each neuron in each layer takes as input all of the outputs from the previous layer only; there are no connections between neurone of the same layer (Figure 3 - An Artificial Neural Network with a single hidden layer). ANNs can feature any number of neurone per layer, different numbers of neurone per layer, and any number of hidden layers. There are trade-offs to each configuration; greater numbers of neurone and layers may be able to approximate non-linear functions to a greater capacity than single layer networks (Bedo, 2014a), though at the cost of added computational complexity and processing time.

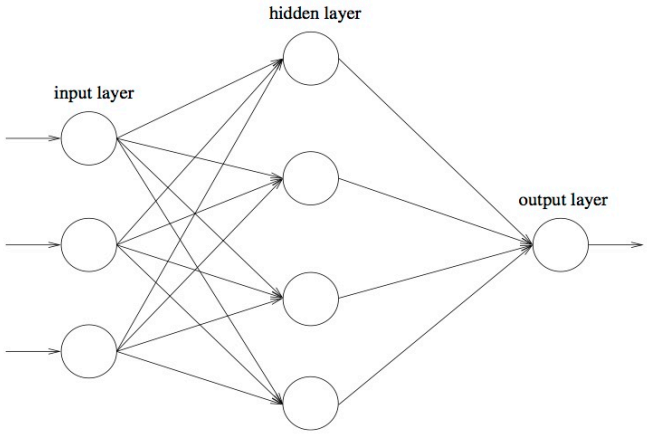

Figure 3 - An Artificial Neural Network with a single hidden layer (Engelbrecht, 2007, p. 7)

The feed-forward neural network proceeds iteratively stepping through each layer from the input and is formalised by Equation 1 (Bedo, 2014a):

$$f_j(x) = \begin{cases} \phi\left(f_{j-1}(x)W_j + b_j\right) & if\ j \geq 1 \\ x & otherwise \end{cases} \quad (1)$$

Equation 1 describes the output (activations) of the first layer as the input values x of the training instance. Activations $f_j(x)$ deeper into the network are produced as follows for the j[th] layer:

1. The previous layer activations $f_{j-1}(x)$ are multiplied by a weight matrix for the current layer W_j as a dot product. This has the effect of both applying weights to inputs and passing them into the transfer function (summation) for each neuron in the current layer.
2. These intermediary signals are summed pairwise with a bias vector b_j for the current layer, producing a second vector of intermediary signals.
3. These final intermediary signals are then each passed into the activation function ϕ to produce the activations for the layer.

A feed-forward neural network supplied with the appropriate weights and biases can act as a function approximation for non-linear relationships between input and output values (Engelbrecht, 2007). However, save for the simplest case, these weights and biases are unknown and require network training through a supervised learning mechanism such as the Back-propagation algorithm.

BACK-PROPAGATION

Discovered by Rumelhart et al. (1988), the back-propagation algorithm advanced research into neural networks by proposing a means for a standard feed-forward neural network to "learn" from prediction error using a concept of gradient descent on the error function. Conceptually, supplying the derivative of the activation with the error signal produced by the previous layer's activation (working backwards from the output layer) generates an error delta which adjusts both the weights and biases in the direction of decreasing error, the gradient descent.

This delta is applied on a per instance basis (in the stochastic case) and, along with a number of optimisations, effectively trains the network by seeking out local error minima across all feature dimensions. However, Hastie et al. (2009) cautions that the error function of the ANN is non-convex, therefore it is important to note that any local error minima discovered by the network are *not* guaranteed to be global.

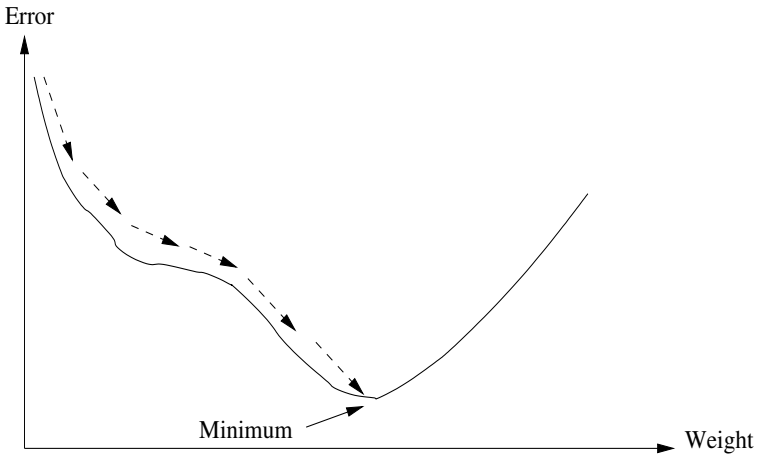

Figure 4 - A 2D plot of the error function with regards to weight, illustrating gradient descent (Engelbrecht, 2007)

The back-propagation algorithm relies on computing the derivative of activations by passing them into the derivative of the activation function itself. The derivative of the activation function (Equation 2) is described in Equation 3:

$$\phi(x) = \tanh(x) \tag{2}$$

$$\phi'(x) = 1 - \tanh^2(x) \tag{3}$$

The back-propagation algorithm can be described in pseudo code (adapted from Bedo, 2014a):

1. Initialise weights W and biases b randomly
2. Until converged
    a. For each training instance
        i. $W_i \leftarrow W_i - \tau\delta$
        ii. $b_i \leftarrow b_i - \tau\delta$

Where "convergence" is described as meeting some stopping criteria (such as maximum number of iterations or steps), W_i is the weight matrix for the current layer, τ is the step size (how far to move in the direction of descending gradient), δ is the delta (error signal), and b_i is the bias vector for the current layer. The delta is calculated using Equation 4 (Bedo, 2014d):

$$\delta_l = \begin{cases} f'_l(x) \otimes (\delta_{l+1} W_{l+1}^T) & if\ l < L \\ (f_L(x) - y) & otherwise \end{cases} \tag{4}$$

The algorithm proper can be illustrated using an adjunct flow diagram (Figure 5); where (A) the error signal δ_l from the last (output) layer is used to update the bias vector (B) and weight matrix (C) of the next-to-last layer. δ_l is then multiplied by the weight matrix (D) to create a new error signal (D). This signal steps the algorithm back another layer and is multiplied by the derived activations (E) to produce a new δ_l where the algorithm begins the next iteration.

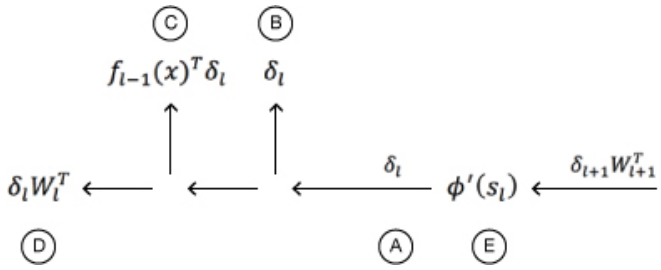

Figure 5 - Adjunct flow diagram of back-propagation (adapted from Bedo, 2014b)

MOMENTUM

A simple optimisation strategy for ANNs is to exploit a previously negotiated step in a given direction for the gradient descent on weight matrices by assuming "momentum" for the error signal being passed back through the network (Bedo, 2014d). Working backwards from the output layer, this is achieved by pairwise addition of the error delta of the previous iteration *t-1* to that of the current *t*, weighted by some proportion $0 < m < 1$ (Equation 5).

$$\delta_W^{(t)} = \delta_W^{(t)} + m\delta_W^{(t-1)} \tag{5}$$

LEARNING RATE DECAY

When conducting stochastic gradient descent, the learning rate parameter refers to the step size in which the model moves to find local minima. The magnitude of this parameter can greatly affect the performance of the model whereby too small a value will prolong convergence and too great a value could overshoot the minimum. There exist a number of optimisation options for this problem (Bedo, 2014d) and chief among these is the implementation of learning rate decay whereby the step size is degraded over time by some parameter related to the current iteration of the model.

For the purposes of this study, we will use a simple learning rate η, which simply decays the step size by a fixed amount α at each iteration t as described in Equation 6 (Bedo, 2014b).

$$\eta_{t+1} = \alpha\eta_t \tag{6}$$

DROPOUT REGULARISATION

One of the principle criticisms of ANNs is their propensity to overfit a solution to input data. This is largely due to the data-driven modification of weights and biases throughout model execution (Hastie et al., 2009). Dropout regularisation is one technique to offset the effects of over-fitting. In this technique, we arbitrarily modify the activations and derived activations from a given layer. Using a probability parameter ρ, we adjust the output of the layer by generating a dropout vector ξ_l, where each element is produced using Equation 7 (Bedo, 2014c). This vector is then applied to both the activations $f_l(x)$ and derived activations $f'_l(x)$ as per Equations 8 and 9.

$$\xi_l = \begin{cases} 0 & with\ probability\ \rho \\ \dfrac{1}{1-\rho} & otherwise \end{cases} \tag{7}$$

$$f_l(x) = \phi(f_{l-1}(x)W_l + b_l) \otimes \xi_l \tag{8}$$

$$f'_l(x) = \phi'(f_{l-1}(x)W_l + b_l) \otimes \xi_l \tag{9}$$

This has the effect of applying a small amount of "brain damage" to the ANN, reducing its dependency on any patterns that may emerge throughout execution as a result of the incoming data and effectively causing a reduction in over-fitting in general. In practice, one normally applies Dropout Regularisation either by assigning a random proportion or a proportion appropriate to the problem through initial parameter trials.

GENETIC ALGORITHM

Considering the high number of possible parameter initialisations available for an ANN and the influence of the data itself being modelled, it is incredibly difficult to determine the optimal configuration on a case-by-case basis (Nguyen, 2014a). To manage this, researchers have looked to evolutionary computation (EC), whose objective is to mimic the natural world through a Darwinian principle of natural selection or "survival of the fittest" (Engelbrecht, 2007; Nguyen, 2014a).

GA is described in terms of "candidate solutions", model instances of varying parameter configurations, which are tried and tested against training data to establish a quantifiable "fitness" of each model instance, given some configuration of parameters. For our purposes, we have selected the performance metric of Root Mean Squared Error (Equation 10), a general-purpose metric that severely punishes large error across the entire data set (Kaggle, 2015).

Specifically, we will compare the predicted ice thickness $\hat{y}_i$ with that of "ground truth" observations y_i from our generated data set. In order to rank candidate solutions in terms of their percentage accuracy, we formalise our fitness score as per Equation 11, which also

produces a fitness score of 0 for solutions off by more than 1 metre (an additional difficulty benchmark).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{10}$$

$$Fitness = 1 - RMSE \tag{11}$$

The fitness of a candidate solution f_i is then evaluated against that of the total population $\sum_j f_j$ to establish a proportional probability of selection p_i for combination (see below), typically through Roulette Wheel Selection (De Jong, 1975), formalised in Equation 12.

$$p_i = \frac{f_i}{\sum_j f_j} \tag{12}$$

The algorithm proceeds until stopping criteria are met through discrete phases, called "generations", which themselves include phases of combination, mutation, and selection/competition (see below). Typical stopping criteria for a GA include a finite number of generations, maximum memory or time usage, or when the reduction in error (or fitness gain) is negligible for the computing power required for a subsequent generation (Safe et al., 2004). Early termination of the GA may lead to a sub-optimal solution and the computing power required to train and test multitudes of models is often the prohibitive factor on large datasets. For simplicity, our GA/ANN model will use a maximum generation count of ten.

The algorithm can be parallelised by distributing the training and testing of disconnected models across compute nodes, reducing their collective solutions together for combination and selection/competition phases (see below). Other possible parallelisation options include programming candidate solutions as agents who select their own partners for combination as they become available (once computation has been completed on another node for instance) (Muhlenbein, 1991), however, this and parallelisation of the GA in general are beyond the scope of this study.

EXPRESSION

Genetic Algorithm approaches rely on a simple principle of expressing each candidate solution in terms of a "genotype" (or chromosome) (Nguyen, 2014a); a static, encoded representation of the parameters used to produce the solution's overall fitness score. In this study we opted for the vector-based encoding strategy, which encodes model parameters as indices on a vector and remains human-readable. Vector representations suffer fewer drawbacks than other encoding strategies, such as binary encoding under which a single bit flip can significantly alter a genotype (Nguyen, 2014a).

The solution encoding for this study takes the following form:

```
[
n_inputs, n_outputs, n_neurone, n_layers,
n_steps, step_size, step_decay,
weight_high, weight_low, bias_high, bias_low,
momentum
]
```

COMBINATION

Combination (or reproduction) is the process by which the fittest candidate solutions from an initial population are "spliced" to create a new candidate solution C* to be trialled, hopefully leveraging the best features of the original parent solutions (Nowostawski & Poli, 1999). In practice, this is implemented simply by selecting an arbitrary index on one parent solution C1 genotype vector as a "cross over" point to exchange with the other parent solution C2. For example:

```
C1 = [1,2,3,4,5]
C2 = [a,b,c,d,e]
C* = [1,2,3,d,e]
```

The resultant genotype vector (the offspring candidate solution) is thereby a combination of two parent solutions.

MUTATION

Further to the processes of combination, the resultant candidate solution is mutated by way of a minor adjustment to one of the values on the genotype. Depending on the problem and data type, this adjustment can be as arbitrary as randomly assigning a new allowable value to the parameter, or as precautionary as adding a small Gaussian perturbation to the value (Nguyen, 2014b). For this study we have opted for the latter, as it allows for exploration within the model space though not at the expense of previously evolved positive traits in the preceding generation.

In practice, mutation is applied in a similar manner to combination, by selecting an arbitrary index on the solution genotype for mutation against a set of mutative functions appropriate for the selected index, such as the addition of Gaussian noise or an integer increment/decrement. For example:

```
C* = [1,2,3,d,e]
C* = [1,2.5,3,d,e]
```

Once mutation has completed for the new candidate solution, it is evaluated against the original training dataset for fitness and added to the total population for the next phase.

SELECTION/COMPETITION

The final phase of the GA approach is to assess the fitness of all candidate solutions, original and generated (as a result of combination/mutation). This process again invokes the Roulette Wheel selection (De Jong, 1975) formalised in Equation 12. This marks the end of the current generation and the model proceeds to the next. As with other GA stages that compute the fitness of candidate solutions that requires a full execution of the model for each instance, fitness calculation is the most CPU intensive portion of the algorithm (Nowostawski & Poli, 1999).

DATA

Three data sets were considered for use in this study:

- Field data have been acquired for Antarctica New Zealand event K063 (2011 and 2013) and have been provided for this study by Gateway Antarctica (PI W. Rack)
- Landsat-8 data is available from the U.S. Geological Survey and was acquired using the USGS EarthExplorer service
- The Terra-SAR-X images were provided through DLR project OCE1592

Data was interpolated, harmonised and cleansed prior to model assimilation.

DATA INTERPOLATION

Given the sparseness of field measurement data, interpolation was applied to generate additional data points with which to train the model. Using QGIS, the Triangular Irregular Network (TIN) interpolation method was chosen, which uses a Delaunay Triangulation to generate the tessellation (QGIS Documentation Team, 2015). This local method was chosen due to its ability to quickly incorporate discontinuity between features (Mitas & Mitasova, 1999), however, we acknowledge that it is one of the least accurate interpolation methods (Mitas & Mitasova, 1999; Nielson, 1993; Renka & Cline, 1984). Points were generated as a 300x300 grid over the triangulation using barycentric interpolation for point values between physical measurements (once for each physical measurement feature) to produce a larger data set for model training. Davis (2010) explains this process as a computation over of the triangular areas A_i formed by an unknown point p within a triangulation of known points.

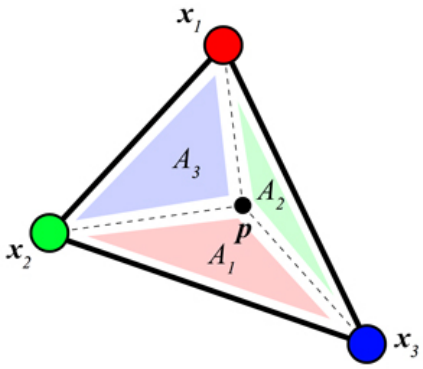

Figure 6 - Barycentric interpolation, using feature values and triangular subareas to compute the value of point *p*. (Davis, 2010)

These areas, which represent a proportional weighting towards a known value (e.g. the size of the triangular area opposite each point represents the tendency towards that feature value), are multiplied by the value of the feature of interest x_i (e.g. snow freeboard), the total of which is then divided by the entire area of the triangle to produce the value of the feature at p. This process is illustrated in Figure 6 and formalised in Equation 13.

$$p = \frac{\sum_{i=1}^{3} A_i x_i}{\sum_{i=1}^{3} A_i} \tag{13}$$

The result of this process produced an interpolated raster layer for each of the physical measurement characteristics, which were added to the training data sets as raster bands (Figure 7, Figure 8).

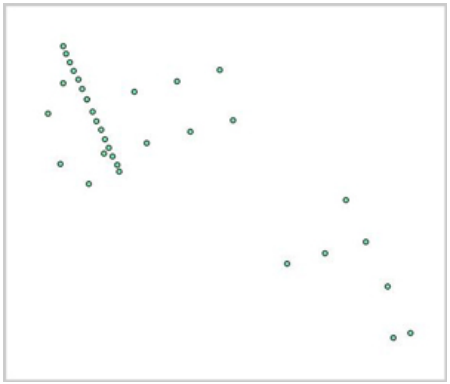

Figure 7 - Raw physical sea ice measurements

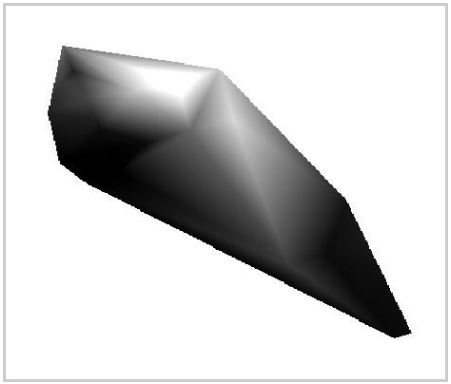

Figure 8 - Sub-ice platelet (SIP) measurements interpolated using TIN (Delaunay) and barycentric interpolation

DATA HARMONISATION

The data sets used in the study come from a variety of sources, stored in different projections, and covering different geographical extents. To harmonise the data, all data sets were reprojected into the "EPSG:3031 Antarctic Polar Stereographic" projection. Masking was used to first clip the resultant data sets to the domain covered by point measurements by producing a polygon masking through a convex hull operation. The data was then further clipped to the SAR imagery, which covers the absolute minimum intersection of all data sets. Figure 9 illustrates a partial sequence of harmonisation operations.

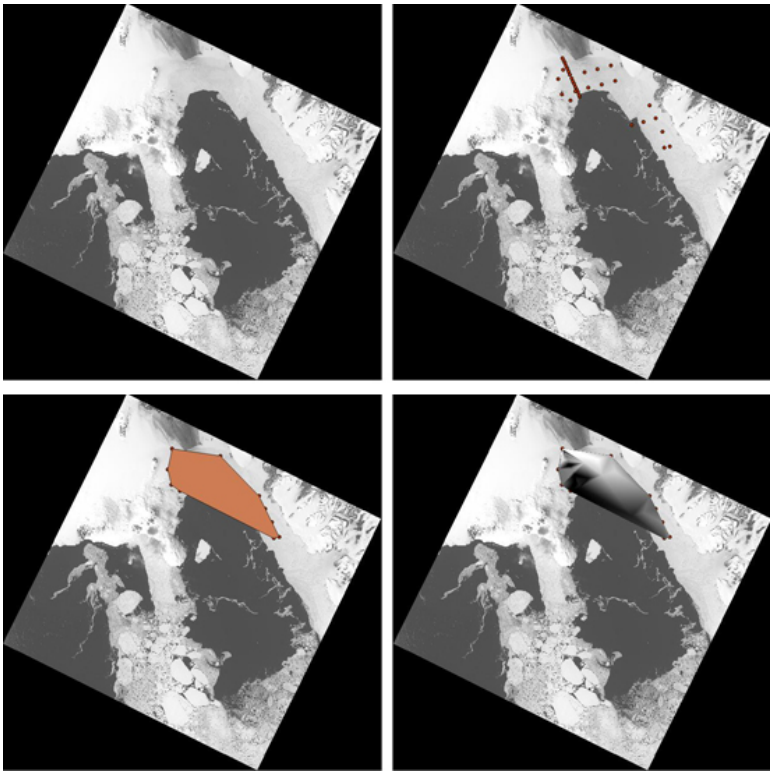

Figure 9 - Partial sequence of data interpolation and masking operations

Data was then converted into a single 17-band raster file for processing (note: bands were merged arbitrarily in software and values were resampled to account for differing resolutions).

Table 1 - Final merged raster band configuration

Band	Original Data
1	Snow Freeboard
2	Ice Freeboard
3	Snow
4	Ice
5	SIP
6	Landsat-8 Band 11
7	Landsat-8 Band 9
8	Landsat-8 Band 8
9	Landsat-8 Band 7
10	Landsat-8 Band 6
11	Landsat-8 Band 5
12	Landsat-8 Band 4
13	Landsat-8 Band 3
14	Landsat-8 Band 2
15	Landsat-8 Band 1
16	Landsat-8 Band 10
17	SAR

A final step taken to harmonise the data was to establish valid transects in code by iterating through each of the data points through all raster bands. Valid transects were determined to be those that featured SAR data, which was the minimum data set for coverage across all data sets. These transects were then enumerated into a flat denormalised form (csv) for model input (Figure 10).

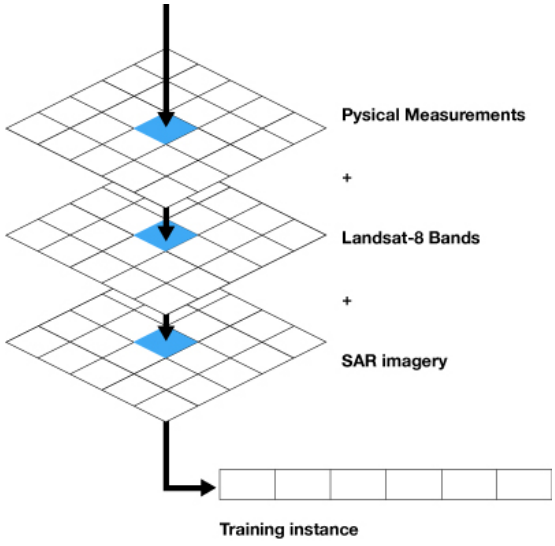

Figure 10 - Transect diagram for instance generation

The resultant data set featured 36,194 training instances, of which 10,000 randomly sampled to train the model and 1,000 were randomly sampled to test the model as mutually exclusive sets.

DATA CLEANSING AND OUTPUT SCALING

Given the presence of corrupt data, or empty cells, it was necessary to further cleanse the data of outliers that would adversely affect model performance. Examples of this include values of non-data placeholders (-9999) and rows with a high concentration of 0 values. The target outputs (sea ice thickness measurements) were also scaled by a factor of 10, allowing the values to reside within the range [0,1], continuous values for approximation by the activation function of the network.

RESULTS

The model was executed using different configurations of data features over ten GA generations. Figures 11-13 below illustrate the most accurate model output. At a glance, the physical measurement model run outperformed both the Landsat-8 and combined data sets, however, it appears that the model suffered from premature culling of high performing candidates in later generations.

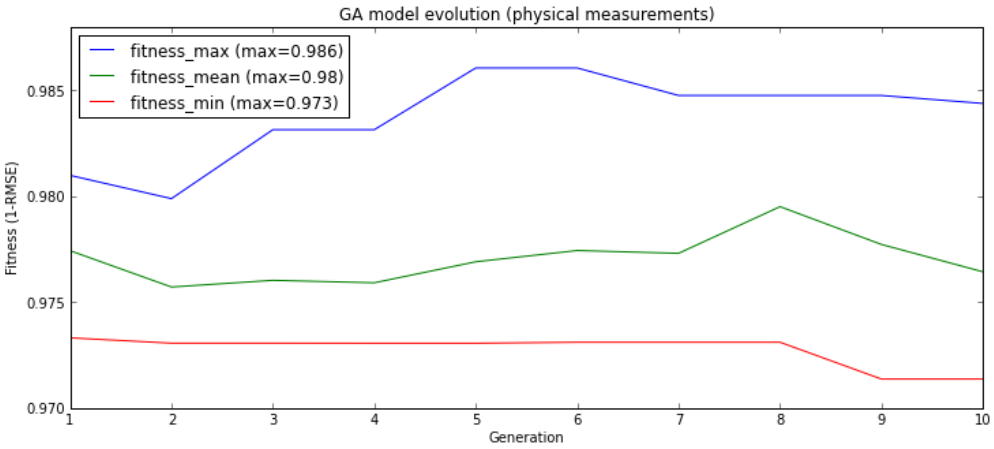

Figure 11 - GA model output using data for physical measurements alone.

This decline in performance is also observed in the Landsat-8 model run, however, the performance is comparable considering the model was trained using remotely sensed data alone.

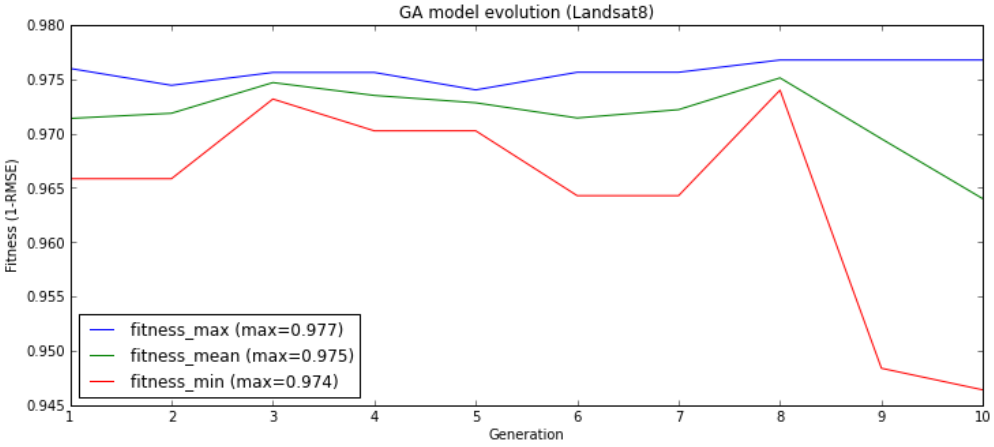

Figure 12 - GA model output using data for Landsat8 bands relevant to sea ice.

The final model run using the combined data set performed better than the Landsat-8 data alone, though not as well as the physical measurements. The combined model also appears to produce a more homogeneous population of candidate solutions.

These observations will be discussed in the following section.

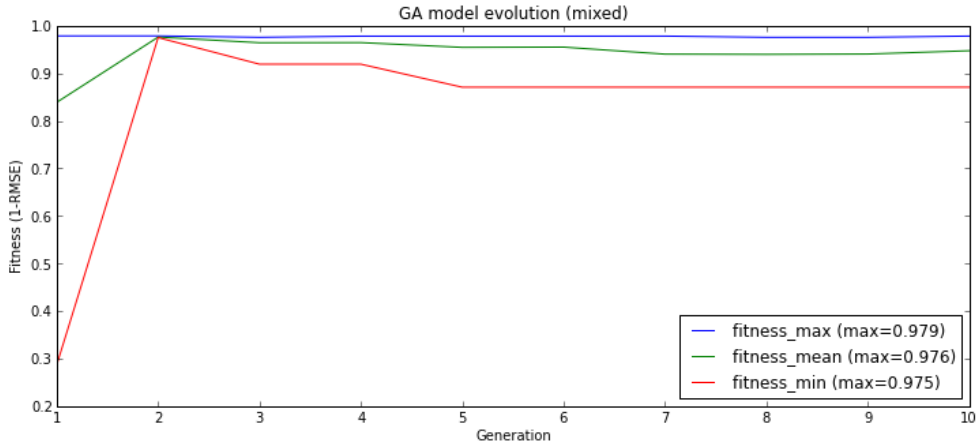

Figure 13 - GA model output using physical observations and Landsat8 bands relevant to sea ice.

DISCUSSION

The first of the model runs trained and tested the model using the full 17-band feature set (16 training features + 1 target feature). Under these conditions, the model failed to model any feature relationships within ten GA generations. Specifically, the output neuron of the network (the thickness prediction) activated at either 1 or -1 (the bounds of the activation function itself), indicating that the features were non-conducive to a meaningful thickness measurement, saturating the activation function. This is likely attributed to the varying magnitude of features, whereby physical measurements were in the range of 0 to 5 whereas the Landsat-8 and SAR imagery were in the realm of tens of thousands. This could possibly have been overcome through additional data manipulation where large inputs could be scaled down to a magnitude comparable to other features (Hastie et al.).

Likewise, model runs using the SAR imagery also underperformed and failed to map any meaningful relationships between input features and ice thickness. We suggest that this is attributed to the nature of the SAR data itself. Hossain et al. (2014) explain the challenges faced by SAR methods in capturing sea ice characteristics in terms of three means of scattering that can occur:

a) Surface scattering – scattering that occurs as the result of uneven surface characteristics
b) Double-bounce scattering – scattering that occurs as the result of an initial reflection (from the surface) encountering another reflective surface (such as a sea ice ridge) before returning to the sensor
c) Volume scattering – scattering that occurs after penetrating the ice and encountering internal structure such as air bubbles or brine pockets

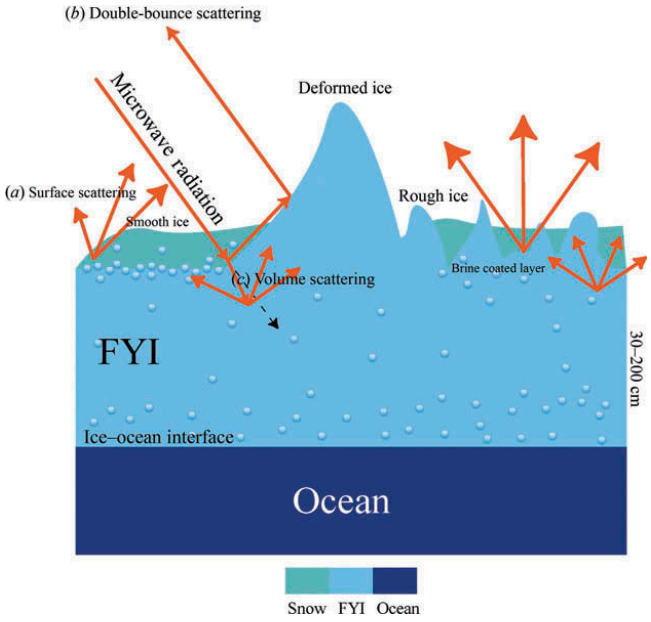

Figure 14 - Methods of scattering (Hossain et al., 2014)

Figure 14 illustrates the scattering mechanisms of SAR over sea ice, which is further complicated by varying ice types such as first year ice (FY) or multi-year (MY) ice (Hossain et al.). Given the fine resolution of the TerraSAR-X data (1m) (DLR, 2009) these effects should be minimised for the SAR data alone, however, interpolation of the field measurements assumes a smooth progression from one point to the next, which negates the resolution accuracy of the SAR data. This demonstrates the known accuracy concerns raised by Mitas and Mitasova (1999); Nielson (1993); Renka and Cline (1984) for the interpolation method chosen and future studies would benefit from more frequent field measurements rather than manufactured data points.

Progressing from initial model runs including all features and SAR imagery, we opted to take a more considered approach, selecting features for three distinct data sets to be trialled:

1. Physical measurements alone
2. LandSat-8 measurements relevant to sea ice detection, and
3. A combination of the two

Figure 11 illustrates the performance of the model over 10 generations using physical measurements alone. In this case, the model performed well, achieving a maximum fitness of 0.986, which equates to an average thickness prediction error ±14cm. In this model we observe the disparity between high-fitness and low-fitness candidate solutions as being among the highest for the targeted model runs (using pre-selected feature sets), which indicates a significant disparity between candidate solutions. Running the model for longer would likely decrease this disparity, however, the main feature of interest is the decline in performance towards the end of the model run. This decline is attributed to highly-performing candidate solutions being culled from the solution population as a result of the Roulette Wheel selection chosen for this study. Although the probability of selection would remain high for these solutions, the nature of randomisation does still allow for the possibility of a high performing solution to be culled. Nguyen (2014b) suggests the use of an "elitism" framework whereby the highest performing solutions are retained regardless of selection probabilities. This could be implemented by simply ranking candidate solutions in order of decreasing fitness and truncating the population back to size for the next iteration. Future studies in this area should consider the application of such a strategy to maximally exploit high performing solutions.

Figure 12 illustrates the performance of the model when using Landsat-8 bands that are relevant for sea ice detection purposes. These bands were selected using information from Schmit et al. (2005), which outlines the operational bands for other sea ice detection algorithms. These band designations have been collated in the following table.

Table 2 – Landsat-8 band designations relevant to sea ice (USGS, 2014)

Band	Wavelength	Details
Band 4	0.64 – 0.67	Red
Band 6	1.57 – 1.65	Short-wave Infrared (SWIR1)
Band 7	2.11 – 2.29	Short-wave Infrared (SWIR2)
Band 10	10.6 – 11.19	Thermal Infrared (TIRS1)
Band 11	11.50 – 12.51	Thermal Infrared (TIRS2)

The Landsat-8 model run did not perform as well as the physical measurements alone, achieving a maximum fitness of 0.977, which equates to an average thickness prediction error of ± 23cm. These numbers are promising considering that they were derived using remotely sensed data alone. However, we note that the initial population of the GA presented a fitness score close to the maximum before the body of the algorithm commenced. This could be the result of over fitting to the data given that the thickness targets were within a fairly narrow range (X - Y) and more data, from different geographical locations could be used to validate the models ability to be generalised across different geographical domains. Furthermore, we note that the Landsat-8 model run also suffers from a variable level of performance with a significant decline towards the end of the model run. Again, this is likely caused by roulette wheel selection used in the survival phase, which prematurely culls high performing models and could be suitably remedied with an "elitism" strategy as suggested by Nguyen (2014b).

The final model run (Figure 13) illustrates the effect of combining both the physical measurements with the Landsat-8 band information. This model run experienced a marked convergence of minimum and maximum fitness among the candidate population by generation 2, suggesting homogeneity of performance between solutions. Although this homogeneity declines and plateaus in subsequent generation, the combined model run demonstrates the contribution of both data sets towards a final solution, decreasing the average error prediction to 0.979 (±21cm) from the Landsat-8 data alone. This illustrates that the combination of the two data sets produces a more homogeneous model population than previous model runs, which depending on the margin for error allowed, may offer a more desirable solution for researchers.

Though these model runs present promising results, they should be taken with caution. Over-fitting of the data is a realistic possibility given the high performance of the models prior to the execution of the main body of the Genetic Algorithm and should be verified with additional data and analysis before any conclusive remarks can be made on the viability of the ANN approach to sea ice thickness in the Antarctic. Furthermore, physical measurements are clearly the most contributing factor in deriving thickness, both outperforming the Landsat-8 model and inherently increasing model performance in the combined approach. However, physical measurements also present the most logistically costly means to collect data for model

assimilation, which may be outweighed by the higher resolution and availability of remotely sensed data at the cost of model performance.

CONCLUSION

This study has investigated the applicability of the Artificial Neural Network model to deriving sea ice thickness from a combination of physical field measurements, Landsat-8 satellite imagery, and Synthetic Aperture Rader (SAR) imagery. Model performance relies on the selection of features for training that are both valid and relevant to the problem domain. Results indicate that the magnitude of input features should be reasonably uniform, and that input scaling may be used to achieve this uniformity. SAR imagery, although providing high resolution, is also sensitive scattering effects as well as to the interpolation technique used for physical measurements in this study (Triangular Irregular Network [TIN] with Delaunay Triangulation and barycentric interpolation). Model performance indicates that physical field measurements are the highest contributor to model performance, outperforming the Landsat-8 model and increasing performance through a combined approach. However, physical measurements also present the higher logistical cost for collection among the data sets studied which presents an argument for acceptable error thresholds using remotely sensed data alone.

The performance of the models in this study suggest that the Artificial Neural Network may be suitable in deriving Antarctic sea ice thickness using a combination of physical field measurements and remotely sensed data. Future studies should consider applying additional optimisations to the Artificial Neural Network and Genetic Algorithm approach used and suggested throughout this study, including the use of an elitism strategy to ensure the highest performing models survive, as well as those optimisations not covered in the scope of this project. Care should also be taken in the interpolation of sparse measurement data, with an emphasis in additional, actual measurements over manufactured data points as a result of inaccurate interpolation. In doing so, future models may also find means to leverage the high resolution of SAR imagery, which was not successfully exploited in this study.

REFERENCES

Alhumaidi, S. M., Jones, W. L., Park, J.-D., & Ferguson, S. M. (1997). A neural network algorithm for sea ice edge classification. *Geoscience and Remote Sensing, IEEE Transactions on, 35*(4), 817-826.

Bedo, J. (2014a). Deep Learning 1. Lecture conducted at University of Melbourne.

Bedo, J. (2014b). Deep Learning 2. Lecture conducted at University of Melbourne.

Bedo, J. (2014c). Deep Learning 3. Lecture conducted at University of Melbourne.

Bedo, J. (2014d). Deep Learning Review. Lecture conducted at University of Melbourne.

Belchansky, G., Alpatsky, I., Eremeev, V., Mordvintsev, I., Platonov, N., & Douglas, D. (2003). *Estimating multiyear sea-ice concentration using passive microwave data and MLP neural networks.* Paper presented at the Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International.

Davis, J. (2010). *Barycentric Interploation.* Retrieved from https://classes.soe.ucsc.edu/cmps160/Fall10/resources/barycentricInterpolation.pdf. Accessed 12 February 2015

De Jong, K. A. (1975). Analysis of the behavior of a class of genetic adaptive systems.

DLR. (2009). *TerraSAR-X - The German Radar Eye in Space.* Retrieved from http://www.dlr.de/eo/en/Portaldata/64/Resources/dokumente/TSX_brosch.pdf. Accessed February 10 2015

El-Diasty, M. K. E.-E. (2003). *Prediction of sea ice concentration using artificial neural networks.* Dissertation/Thesis, ProQuest, UMI Dissertations Publishing. Retrieved from http://search.proquest.com.ezproxy.canterbury.ac.nz/docview/305214252

El-Rabbany, A., Auda, G., & Abdelazim, S. (2002). Predicting sea ice conditions using neural networks. *Journal of Navigation, 55*(01), 137-143.

Engelbrecht, A. P. (2007). *Computational intelligence: an introduction*: John Wiley & Sons.

Hassoun, M. H. (1995). *Fundamentals of artificial neural networks*: MIT press.

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., & Tibshirani, R. (2009). *The elements of statistical learning* (Vol. 2): Springer.

Hossain, M., Yackel, J., Dabboor, M., & Fuller, M. C. (2014). Application of a three-component scattering model over snow-covered first-year sea ice using polarimetric C-band SAR data. *International Journal of Remote Sensing, 35*(5), 1786-1803.

Kaggle. (2015). *Root Mean Squared Error.* Retrieved from https://www.kaggle.com/wiki/RootMeanSquaredError. Accessed 29 January 2015

Kendall, G. (2001). *Neural Networks.* Retrieved from http://www.cs.nott.ac.uk/~gxk/courses/g5aiai/006neuralnetworks/neural-networks.htm. Accessed 30 January 2015

Kurtz, N. T., Farrell, S. L., Studinger, M., Galin, N., Harbeck, J. P., Lindsay, R., . . . Sonntag, J. G. (2013). Sea ice thickness, freeboard, and snow depth products from Operation IceBridge airborne data. *Cryosphere, 7*(4), 1035-1056. doi: 10.5194/tc-7-1035-2013

Kwok, R. (2011). Satellite remote sensing of sea-ice thickness and kinematics: A review. *Journal of Glaciology, 56*(200), 1129-1140.

Lemke, P., Ren, J., Alley, R. B., Allison, I., Carrasco, J., Flato, G., . . . Thomas, R. H. (2007). Observations: Changes in Snow, Ice and Frozen Ground, In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge, United Kingdom and New York, NY, USA.: Cambridge University Press.

Lin, H., & Yang, L. (2012, 2012). *A hybrid neural network model for sea ice thickness forecasting.* Paper presented at the 8th International Conference on Natural Computation.

Massom, R. A., Eicken, H., Haas, C., Jeffries, M. O., Drinkwater, M. R., Sturm, M., . . . Allison, I. (2001). Snow on Antarctic sea ice. *Reviews of Geophysics, 39*(3), 413-445. doi: 10.1029/2000RG000085

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics, 5*(4), 115-133.

Minsky, M., & Papert, S. (1969). Perceptron: an introduction to computational geometry. *The MIT Press, Cambridge, expanded edition, 19*, 88.

Mitas, L., & Mitasova, H. (1999). Spatial interpolation. *Geographical information systems: principles, techniques, management and applications, 1*, 481-492.

Muhlenbein, H. (1991). *Evolution in time and space-the parallel genetic algorithm.* Paper presented at the Foundations of genetic algorithms.

Nguyen, V. (2014a). Evolutionary Computation 1. Lecture conducted at University of Melbourne.

Nguyen, V. (2014b). Evolutionary Computation 2. Lecture conducted at University of Melbourne.

Nielson, G. M. (1993). Scattered data modeling. *IEEE Computer Graphics and Applications, 13*(1), 60-70.

Nowostawski, M., & Poli, R. (1999). *Parallel genetic algorithm taxonomy.* Paper presented at the Knowledge-Based Intelligent Information Engineering Systems, 1999. Third International Conference.

QGIS Documentation Team. (2015). *Spatial Analysis (Interpolation).* Retrieved from http://docs.qgis.org/2.2/en/docs/gentle_gis_introduction/spatial_analysis_interpolation.html. Accessed 30 January 2015

Renka, R., & Cline, A. (1984). A triangle-based C1 interpolation method. *Rocky Mountain Journal*.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review, 65*(6), 386.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1988). Learning representations by back-propagating errors. *Cognitive modeling, 5*.

Safe, M., Carballido, J., Ponzoni, I., & Brignole, N. (2004). On Stopping Criteria for Genetic Algorithms. In A. C. Bazzan & S. Labidi (Eds.), *Advances in Artificial Intelligence – SBIA 2004* (Vol. 3171, pp. 405-413): Springer Berlin Heidelberg.

Schmit, T. J., Gunshor, M. M., Menzel, W. P., Gurka, J. J., Li, J., & Bachmeier, A. S. (2005). INTRODUCING THE NEXT-GENERATION ADVANCED BASELINE IMAGER ON GOES-R. *Bulletin of the American Meteorological Society, 86*(8), 1079-1096. doi: 10.1175/BAMS-86-8-1079

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., . . . Miller, H. (2007). Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom and New York, NY, USA.: Cambridge University Press.

USGS. (2014). *Frequently Asked Questions about the Landsat Missions.* Retrieved from http://landsat.usgs.gov/band_designations_landsat_satellites.php. Accessed 1 February 2015

Wadhams, P. (2000). *The Frozen Oceans: Ice in the Ocean*. Amsterdam, The Netherlands: Overseas Publishers Association.

Xie, H., Ackley, S. F., Yi, D., Zwally, H. J., Wagner, P., Weissling, B., . . . Ye, K. (2011). Sea-ice thickness distribution of the Bellingshausen Sea from surface measurements and ICESat altimetry. *Deep-Sea Research Part II, 58*(9), 1039-1051. doi: 10.1016/j.dsr2.2010.10.038

Yap, H. J., Lim, W. K., Ewe, H. T., & Chuah, H. T. (2009). Passive Microwave Remote Sensing for Sea Ice Thickness Retrieval Using Neural Network and Genetic Algorithm. *PIERS Online, 5*(2), 191-195.